FPS FORENSIC PSYCHOLOGICAL SERVICES
Supporting your Criminal and Civil Justice Needs

Middlesex University London

The Justice Data Lab

Synthesis and Review of Findings

September, 2016

Joanna R Adler and Mark Coulson
Department of Psychology, School of Science and Technology, Middlesex University

www.mdx.ac.uk/fps
fps@mdx.ac.uk

Justice Data Lab: Synthesis and Review of Findings

Professor Joanna R. Adler and Dr Mark Coulson

Contents

1 Executive Summary

This paper provides a brief review of the operation of the Justice Data Lab (JDL), setting its work in the context of literature regarding effective rehabilitation of people who have offended. At the paper's core is a request to begin to synthesise the findings of the JDL with a view to considering the ways in which the reports and findings from the JDL may be contributing to the wider evidence base about intervention and management of people who have offended.

1.1 Summary of Findings

1. It is hard to discern many trends given the relatively limited uptake of the service and challenges faced in aggregating the published data;
2. Nonetheless, there is positive evidence that the JDL has made a solid start to its operations;
3. The JDL has been generally well received by those who have used it;
4. It is encouraging to note that so many positive outcomes have been demonstrated across different interventions and sectors;
5. However, the majority of findings have been labelled as inconclusive, even when change has been observed; this has led to uncertainty from JDL clients about how to use such findings;
6. Whether a small or large cohort was put through the analysis, the magnitude of change observed did not vary significantly. This is a positive finding and indicates both that change *can* be observed and that it can be discerned despite statistical 'noise' in the model.
    a. It is not surprising that with relatively modest effect sizes, the changes observed fail to reach statistical significance when derived from small user cohorts;
7. Despite the caveats above, we can conclude that educational interventions (offered by the Prisoners' Education Trust) are repeatedly demonstrated as being effective. There are also some positive findings arising from employability/employment initiatives although there is more variability in outcome here, warranting further investigation.

1.2 Recommendations

Wherever possible, use natural language to summarise findings. We commend recent changes that start to make effect sizes more prominent however we believe that more accessible language would improve the reports still further and suggest additional strategies in this regard.

Increase uptake and engagement. General awareness needs to be raised and if possible, routes should be found to enable smaller provider organisations to collaborate in drawing on the model. However, it is acknowledged that this may be difficult in practice due to the commissioning and implementation contexts in which different, potential competing, organisations work. Concomitantly, organisations should be better guided as to when the JDL measures may be inappropriate.

Provide more support for using the findings. Advice to potential service users could provide more examples of ways in which to use the JDL reports, for example case studies of previous presentations made to Board of Trustees or commissioners may be useful. Better support for use of the findings would also help enhance engagement.

Set up a means to retain and make redacted uploaded data available. We concur with suggestions to retain data and agree that secondary analyses could be invaluable to both academic and policy debates. We note that possible ways in which to archive data and how far they should be made available are already under consideration by the JDL.

Collate more information on intervention practices. In particular, information about intervention frequency, duration and intensity would be useful for future meta-analyses.

2 Introduction

This paper provides a brief review of the operation of the Justice Data Lab (JDL), setting its work in the context of literature regarding effective rehabilitation of people who have offended. It forms part of a wider review, commissioned by NPC that is being led by Professor Fergus Lyon, of CEEDR, Middlesex University. This report will not consider the JDL's economic sustainability, nor its applications to other contexts as both will be considered elsewhere in the independent review. Rather, this paper is centred upon a request to begin to synthesise the findings of the JDL with a view to considering the ways in which the reports and findings from the JDL may be contributing to the wider evidence base about intervention and management of people who have offended. This review is also designed to suggest ways to move forward and to consider uses of such evidence as the JDL becomes an embedded part of the evaluation landscape.

2.1 Authors' Note

The authors of this paper are members of the JDL Expert Panel and we are grateful for the other panel members' input to the first draft. All materials drawn on within this paper are publically available. All views expressed are those of the authors.

3 Background to the Review: Assessing the outcomes and impacts of criminal justice interventions

Forty years ago, Robert Martinson wrote a review of prison rehabilitation programmes and set up the "What Works" question to explore their effectiveness. His initial reading of the limited available evidence led him to knock down the "What Works" question and to conclude that nothing in prison could rehabilitate offenders. However, the initial paper[1] was an early publication from a large scale project[2] that was subsequently more nuanced in its conclusions. Those conclusions can be summarised as: offender rehabilitation programmes were not working; or, they may have been working but the evaluation methodology did not show that they were effective; or, that the programmes were not being given opportunities to work as they were neither funded nor implemented properly. The field has moved on enormously since then and researchers have found adult and youth justice interventions that do seem to be effective. That is, interventions have been demonstrated to have an impact on recidivism or other relevant indicators of improvement such as reduced substance misuse, or sustained employment. The themes of rigour of evaluation and quality of programme implementation are as important now as ever and will be returned to below.

[1] Martinson, 1974
[2] Lipton, Martinson and Wilks, 1975

For now, we summarise briefly the state of knowledge on what works in rehabilitating people who have offended. There are a number of reviews considering how best to manage adults and young people who have offended, how to minimise the likelihood that they will reoffend and how to prevent them offending in the first place. These reviews are mainly American[3] but there are also some relevant European[4] syntheses of findings. A common conclusion that can be drawn is that rehabilitative and therapeutic approaches[5], often with a cognitive behavioural focus[6], work better than more punitive or surveillance programmes[7] and that early intervention, or prevention has the greatest impact and is most cost effective[8]. Restorative approaches have been widely seen as effective within adult populations but the efficacy of such approaches is less clear with young people[9] whereas educational interventions seem to produce strong outcomes in both domains[10]. When working with young people, variations in intervention outcomes have related to matters such as: rigour of evaluation and generalizability of findings[11]; offender characteristics[12]; the impact of staff practice[13]; and judicial system philosophy/approach[14], the importance of their needs as a young person first and offender second, has also been emphasised[15] and a repeated conclusion is that programmes must be implemented to appropriate service users, in ways to maximise therapeutic alliance[16]. A salutary finding also worth noting is that in American settings, diversion away from the criminal justice system, with no intervention at all, could have been more effective than any traditional court sanctions[17].

Over recent years, meta-analysis has been seen as the preeminent way to summarise findings from a range of studies that are broadly related to the same topic. Meta-analysis will also usually form part of a systematic review compiled under, for example, a Cochrane protocol[18]. The principles of rigorous evaluation have been increasingly adopted in making policy decisions and can be seen in projects such as the Mayor of London's youth and evidence hub, 5 levels of evidence[19]. This is itself derived out of an American model that

[3] Aos and Drake, 2013 (see also WSIPP below); Schwalbe et al., 2012
[4] Koehler, Lösel et al., 2013; Redondo et al. 1999
[5] Petrosino et al., 2010
[6] Garrido and Morales, 2007
[7] Petrosino et al., 2013
[8] Ross et al., 2011; Small et al. 2005
[9] Strang et al., 2013
[10] Davis et al., 2013 (for adults); Miller et al., 2015 (for youth)
[11] Mears et al. 2011
[12] Lipsey et al., 2000
[13] Dowden and Andrews, 2004
[14] Koehler, Hamilton et al., 2013
[15] Haines et al., 2013; Adler et al. 2016
[16] England, 2009; Larkins & Wainwright, 2013
[17] Smith et al., 2004
[18] See for example: http://www.cochrane.org/what-is-cochrane-evidence accessed 22-6-16
[19] http://project-oracle.com/support/youth-organisations/validation accessed 23-6-16

rates evidence and that is now widely adopted in European policy making. In the National Institute of Justice (USA) model (NIJ), evidence of an intervention's efficacy is assessed as being either "effective", "promising" or having "no (demonstrable) effect"[20]. Perhaps the best known exponents of this approach are the Washington State Institute for Public Policy (WSIPP). Following State legislation to evaluate efficacy of interventions, the WSIPP team has been compiling efficacy studies, conducting meta-analyses and considering costs and benefits of public policy interventions. For the purposes of the current paper, key information produced by WSIPP includes summary tables of adult and youth justice interventions. Tables A1 and A2 in Appendix 1 show the most recent overviews of programme evaluations summarised by WSIPP and their likelihood of generating positive returns on investment[21]. As can be seen, several programmes seem to be highly effective (100%) whilst others produced a negative return on investment and may even have increased offending overall, (such as "scared straight" at 4% in Table A1).

In 2012, the then available WSIPP data were reanalysed with costings for England and Wales[22]. At a similar time, the Ministry of Justice (MoJ) produced a compendium of justice statistics and analysis (also updated the following year)[23]. In the compendium, MoJ tested the outcomes of different court sanctions against one another (e.g. how would community sentence outcomes compare with those from short prison sentences, if the offender related variables were otherwise similar)? Whilst useful to assess sentencing outcomes, this approach could not indicate what it was about any one sanction that was, or was not, effective. Alongside this, Transforming Rehabilitation was coming into full effect with 21 community rehabilitation consortia (CRC) contracted by February 2015[24] following on from pilots of payment by results at HMPs Doncaster and Peterborough. As the private and voluntary sectors increasingly joined the public sector in the management of people who have offended, it has become of paramount policy importance that efficacy can be tested, particularly in regards to potential reductions in recidivism rates.

4 Why the Justice Data Lab?

As outlined above, there is a growing body of evidence about what works in community and custodial settings to rehabilitate offenders (young and adult). However, it is not clear that because a programme is more effective than alternatives in one country, it will be more

[20] http://www.crimesolutions.gov/about_starttofinish.aspx accessed 23-6-16
[21] http://www.wsipp.wa.gov/BenefitCost?topicId=1 accessed 22-6-16
[22] The Social Research Unit, 2012; 2013
[23] Ministry of Justice 2012, 2013
[24] http://www.clinks.org/criminal-justice-transforming-rehabilitation/what-transforming-rehabilitation

effective than the “treatment as usual” in another country[25]. Also, it is unclear if a programme developed for use with one client group, such as medium risk, adult, male offenders, will work with another client group. Relatedly, it should be noted that the most rigorous evaluations of outcome, may say little about process, i.e. the characteristics of implementation that contribute to whether a programme is, or is not working well. Further, when considering review evidence based on meta-analysis, it is worth considering that although statistically strong, meta-analysis is deliberately limited in scope because of the need to sift evidence according to the rigour of the initial evaluation. A meta-analysis will typically exclude a much larger proportion of potential studies than that included, not necessarily because of a problem with the interventions under consideration but because of a failure of the evaluation of that intervention to meet the necessary statistical thresholds for inclusion. Thus, many interventions will not be considered, not because they have been shown to fail, but because they have not been tested in ways that would enable their inclusion in such reviews.

For the voluntary sector, and small private sector providers, accessing comparator data is a particular problem. They may not have the throughput of service users to be able to construct a randomised control trial, nor may they have the staff expertise or resources to run, analyse or interpret such research. Yet without independent assessment, interventions cannot be tested to see whether they have desirable outcomes and impacts. In policy and commissioning terms, without evaluation, they cannot demonstrate when they are ‘promising’ or ‘effective’, nor of course, if they are ineffective. The Justice Data Lab (JDL) was set up precisely to fill this gap.

The primary purpose of the JDL was “to provide a national system for accessing offender data”[26]. It was intended to provide an accessible alternative to the randomised control trial, one that would enable meaningful comparison in a model that could be replicated and consistently adopted for use in evaluating diverse providers of multi-faceted, offender interventions. In designing exactly how to take public sector provider data and compare it against national norms, the decision was taken early on that propensity score matching would provide a route to enable the creation of intervention and comparator groups that would come as close as possible to a randomised control trial (RCT), without having to bear the resource and time loadings such a trial necessitates and that it would enable comparisons where an RCT was neither feasible nor appropriate.

[25] Adler et al. 2016 for further discussion of this point.
[26] Rickey & Pritchard, 2012.

In using the JDL, providers to the public sector need to be confident that their sensitive data will be kept confidential and that they can trust that the findings will be rigorous, and acceptable to policy audiences as well as to their own boards of management and, or, trustees. Also, they need to be able to provide data to the JDL in ways that are not overly cumbersome to the organisation providing data, yet usable to the JDL staff. The model of analysis adopted by the JDL, ways to obtain data and reporting protocols have all evolved since its inception in 2013 but core principles are well summarised in the User Journey Document available from the JDL and in the report of the pilot year[27]. In Feb, 2016, the JDL published a response to a review of methodology, this will be returned to in the section of this paper "Refining the Methodology".

For now, we summarise the JDL process thus: The organisation wishing to submit its interventions/programmes for analysis provides data on a cohort of service users to the JDL, this we consider as an "intervention group". These data are kept securely for the duration of analysis and unless otherwise agreed, are destroyed after the analysis has been completed. The intervention group's data provided are considered in terms of each person's likelihood to reoffend. This uses a "propensity" model, where the risk of reoffending is considered in terms of predicted outcomes, via "treatment as usual" within community or custodial settings. Each service user is considered in terms of his or her pre-assessed risk propensity. The service users are then compared against multiple other people who have offended and been assessed with similar propensities (the "comparison group"). This is known as propensity score matching and allows for comparison of the intervention's service users against similar others from nationally held data. The specific form of propensity matching adopted by JDL is beyond the scope of this paper, although it should be noted that alternatives were raised in its review of methodology[28]. In the current model, intervention group individuals are matched with as many comparison group participants as fit the specified propensity score, resulting in comparison groups that are much larger (perhaps several orders of magnitude larger) than intervention groups. Note that in both groups (the intervention or comparison), there is no assessment made of which other programmes or interventions the providers' client groups may have engaged with; the model assumes that it is testing one intervention against treatment as usual, where such usual "treatment" might involve multiple other potential interventions.

[27] Ministry of Justice, 2014a & b; 2015a
[28] Ministry of Justice, 2016a

5 Synthesis of Key JDL Findings

The May 2016 JDL publication summary contains data from 132 individual reports published since the scheme's inception in 2013[29]. Reports are classified by Sector, Intervention type, and Intervention group size. This framework is outlined in table 1. Results are further classified in terms of whether they show a reduction in offending, an increase in offending, or are inconclusive at this stage. These reductions/increases are measured in three ways; proven one-year reoffending rate (all reports), one year frequency of offending (118/132 reports), and time to reoffend (25/132 reports). The most recent analyses of reoffending are tiered by severity of offence, have only been performed for 6 reports, and are not considered further here.

Table 1. Descriptive framework for JDL reports.

Sectors Evaluated			
Educational Institution	Private Organisation	Public Organisation	Voluntary & Community
Intervention types			
Accommodation	Arts	Education/Learning	Employment
Health and Wellbeing	Mentoring	Multi-Purpose	Problem solving
Relationship Building	Restorative Justice	Substance Misuse	Youth Intervention
Intervention group sizes			
30-100	101-200	201-300	301-400
401-500	501+		

The synthesis of key findings is based upon published reports of JDL analyses. Before presenting the synthesis, several caveats and assumptions should be noted. These all arise from the ways in which JDL reports are published and are outlined in Appendix 2. The two key assumptions are that:

1. National comparison samples are superior to regional comparison groups as they result in larger matched samples, and correspondingly greater statistical power.

2. Later requests are more reliable indicators of an intervention's effect than first requests. This is because even if no more clients have gone through an intervention

[29] Ministry of Justice, 2016b & c

on re-analysis, the dynamic and ongoing nature of the JDL model utilised means that more recent results are likely to be more accurate estimates of interventions' efficacy.

The result of applying these two assumptions is a reduction in the number of reports to be considered from 132 to 97. Frequencies by sector, intervention and size are provided in table 2 below. It should be noted here that there has been no systematic removal of any particular sector in conducting aggregate analyses. Rather, we have used only the national versions of comparisons conducted, not regional ones and, where the same intervention/programme has been assessed more than once, we have used the most recent report generated. It should also be noted that any one organisation may run more than one intervention, or that the same intervention may be delivered to different client groups (e.g. young men and adult women). Multiple client groups and/or multiple interventions provided by one provider mean that these 97 reports were produced for 38 JDL service users/provider organisations. A full list of these organisations is provided in Appendix 2.

Table 2: Number of reports by Sector, Intervention, & Intervention group Size

Sector Type	**Intervention group Size**						
Intervention Type	**30-100**	**101-200**	**201-300**	**301-400**	**401-500**	**501+**	**Grand Total**
Educational Institution	**0**	**1**	**1**	**1**	**0**	**2**	**5**
Employment		1	1	1		2	5
Private	**2**	**4**	**11**	**3**	**1**		**21**
Employment	2	4	11	3	1		21
Public	**10**	**3**	**6**	**2**	**2**	**4**	**27**
Accommodation	3	1					4
Employment	2	2	6	2	2	4	18
Mentoring	3						3
Problem Solving	1						1
Youth Intervention	1						1
VCS	**18**	**3**	**5**	**5**	**4**	**9**	**44**
Accommodation	3	1	1	3	1	1	10
Arts	1		1				2
Education / Learning				2	1	5	8
Employment	2		1		2	2	7
Health and Wellbeing	1						1
Mentoring	4		1			1	6
Multi-Purpose	1						1
Relationship Building	3	1					4
Restorative Justice		1					1
Substance Misuse	1		1				2
Youth Intervention	2						2
Grand Total	**30**	**11**	**23**	**11**	**7**	**15**	**97**

The richness of the classification system (4 sectors, 12 interventions, 6 size categories) results in a framework with more cells (288) than data points (97). The table is therefore sparsely populated, and will remain so for some time to come, making discerning patterns problematic. One approach is to remove each dimension in turn, collapsing across its levels (e.g. to consider all of any one type of intervention, irrespective of sample size; or to show all of one sample size, irrespective of intervention type). The results of this approach are outlined in tables 3 and 4.

Table 3: Number of reports by Sector and Intervention

	Sector				
Intervention Type	**Educational**	**Private**	**Public**	**VCS**	**Grand Total**
Accommodation			4	10	**14**
Arts				2	**2**
Education / Learning				8	**8**
Employment	5	21	18	7	**51**
Health and Wellbeing				1	**1**
Mentoring			3	6	**9**
Multi-Purpose				1	**1**
Problem Solving			1		**1**
Relationship Building				4	**4**
Restorative Justice				1	**1**
Substance Misuse				2	**2**
Youth Intervention			1	2	**3**
Grand Total	**5**	**21**	**27**	**44**	**97**

There is considerable clustering of interventions within sectors, with two sectors (educational institutions and private sector organisations) only focussing on a single type of intervention (employment in both cases). While both public and voluntary sectors have more varied portfolios, there are few, or no, examples of many intervention categories within any one sector. Employment interventions are more frequently evaluated than all other intervention types combined and 45% of all reports were conducted for the voluntary sector.

Table 4: Number of reports by Sector and Intervention group Size

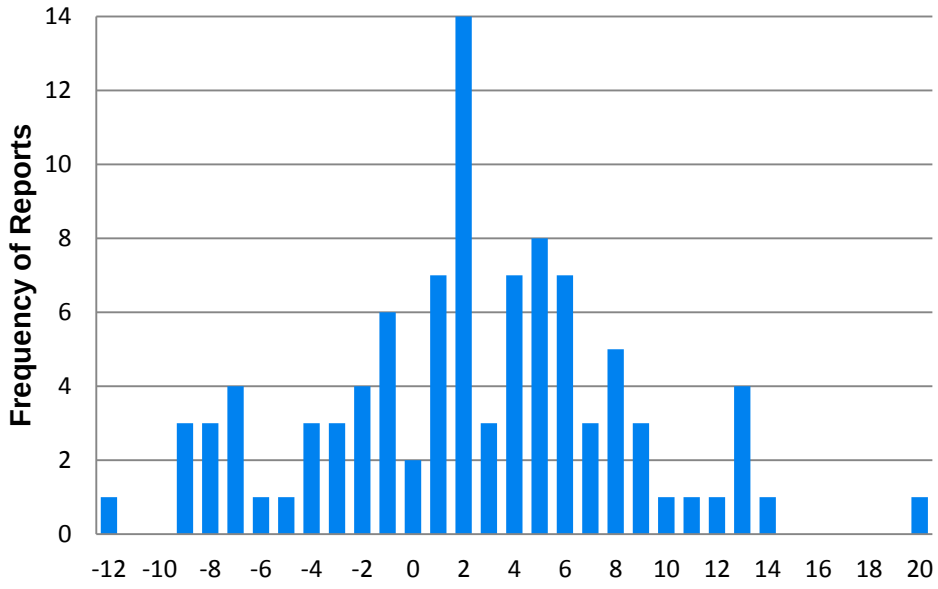
	Intervention group Size						
Sector	**30-100**	**101-200**	**201-300**	**301-400**	**401-500**	**501+**	**Grand Total**
Educational Institution		1	1	1		2	**5**
Private	2	4	11	3	1		**21**
Public	10	3	6	2	2	4	**27**
VCS	18	3	5	5	4	9	**44**
Grand Total	**30**	**11**	**23**	**11**	**7**	**15**	**97**

Each sector presents a unique pattern in terms of intervention group size evaluated, with no clear trends emerging. It may be worth noting however, that 64/97 (66%) of evaluations, were conducted on intervention group sizes of between 30 and 300. This raises the question therefore whether sample size may have an impact on outcomes found.

5.1 Effect Size

The mean effect size across all analysis conducted is 2.4 percentage points. The median effect size is 2 percentage points. Figure 1 shows the distribution of effect sizes. It demonstrates an emerging normal distribution around the mode of 2 percentage points with very few interventions demonstrating changes of more than 10%, in either direction (two of which are clearly outliers). These findings are useful because they remind us of the average effect size we might expect for rehabilitation services; and which should be reflected in practitioners', providers' and service commissioners' expectations.

Figure 1. Effect size (% point) by frequency of Justice Data Lab reports



NB negative=increase in reoffending

Base: 97 Justice Data Lab reports published 2013 – 2016

5.2 Sample Size

There is no evidence that larger samples result in larger effect sizes[30] however, the likelihood of discerning an effect (irrespective of size) is likely to increase as sample sizes increase. For one year reoffending, significant decreases in reoffending are associated with larger sample sizes (average size 747 participants) than inconclusive results (average size 214)[31], and this pattern is repeated for frequency of reoffending[32]. Data for time to reoffending were not analysed due to the low number of available analyses of this type. For both one year reoffending and frequency of reoffending, the sample sizes for the very few interventions showing increased offending (i.e. "not working") were statistically indistinguishable from the other groups.

The distribution of decreases, increases and inconclusive results for both one year reoffending and frequency of reoffending by sample size are outlined in tables 5 and 6.

Table 5: Outcomes on One Year Reoffending Rate, by Sample Size

	Intervention group Size						
Reoffending Rate	**30-100**	**101-200**	**201-300**	**301-400**	**401-500**	**501 +**	**Grand Total**
Decrease	4	2	4	6	4	8	28
Inconclusive	25	9	17	4	3	6	64
Increase	1		2	1		1	5
Proportion of decreases	**13%**	**18%**	**17%**	**55%**	**57%**	**53%**	

[30] One way ANOVA on percentage change for one year proven reoffending, with five sample sizes, $F(4,90) = 1.02$, $p=.400$, $\eta_p^2=.04$).

[31] One way ANOVA with change in one year reoffending (decrease, increase, inconclusive) as fixed factor and treatment group size as the dependent variable, $F(2, 94)=6.24$, $p=.003$, $\eta_p^2=.12$).

[32] One way ANOVA with change in frequency of reoffending (decrease, increase, inconclusive) as fixed factor and treatment group size as the dependent variable, $F(2, 82) = 4.51$, $p=.014$, $\eta_p^2=.10$).

Table 6: Outcomes on One Year Reoffending Frequency, by Sample Size

Reoffending Frequency	Intervention group Size 30-100	101-200	201-300	301-400	401-500	501+	Grand Total
Decrease	4	2	5	7	3	7	28
Inconclusive	17	7	17	3	4	7	55
Increase				1		1	2
Proportion of decreases	**19%**	**22%**	**23%**	**63%**	**43%**	**47%**	

What works in reducing proven one-year reoffending?

There appears to be something of a step in both distributions, with the probability of detecting a significant decrease in reoffending jumping from around 20% for sample sizes of 300 or less, to above 55% for samples larger than 300. This means that for organisations with small sample sizes, their chances of gaining an inconclusive finding are much higher than for those organisations with larger cohorts of service users. As already noted however, this is mainly a manifestation of the sample sizes themselves and their relationship to statistical significance. It is not necessarily something related to the efficacy (or effect size) of the interventions per se.

As the effect sizes do not vary with sample size, we have also aggregated the data across sample sizes to begin to consider “what works” in more detail. In subsequent analyses, we consider all three outcome measures (proven one year reoffending (PR); one year frequency of reoffending (FR);and time to reoffend (TR)) and examine the number of reports which result in decreases, increases, and inconclusive results. We do not report percentages within tables as the very small number of reports in many cells would make this a particularly uninformative, potentially misleading, presentation of the data. For readers’ convenience however, we do use percentages in “what works” discussions below, here too, the low numbers involved should be kept in mind.

5.2.1 Effects on Proven One-Year Reoffending (PR: 97 reports)

Within Tables 7 and 8, the three numbers in each cell refer to the number of analyses which showed a significant decrease in reoffending/a significant increase in reoffending/ were inconclusive. In each cell, we have presented the three figures consistently in this order: decrease/increase/inconclusive.

Table 7: Results for One-Year Proven Reoffending by Intervention & Sector

Intervention	Sector				
	Educational	Private	Public	VCS	Total
Accommodation			1/0/3	3/3/4	4/3/7
Arts				0/0/2	0/0/2
Education / Learning				8/0/0	8/0/0
Employment	1/0/4	5/2/14	5/0/13	3/0/4	14/2/35
Health and Wellbeing				0/0/1	0/0/1
Mentoring			0/0/3	1/0/5	1/0/8
Multi-Purpose				0/0/1	0/0/1
Problem Solving			0/0/1		0/0/1
Relationship Building				0/0/4	0/0/4
Restorative Justice				0/0/1	0/0/1
Substance Misuse				0/0/2	0/0/2
Youth Intervention			1/0/0	0/0/2	1/0/2
TOTAL	1/0/4	5/2/14	7/0/20	15/3/26	28/5/64

5.2.2 What works in Reducing One-Year Proven Reoffending?

Around 29% of interventions (28/97) have been shown to reduce PR. These lie predominantly within Employment, but the difference between the success rate in Employment (14/51) and the overall success rate (28/97) is almost identical. Success rates within Sectors are higher for Voluntary and Community Services than for Private and Public organisations (there are too few observations from Educational Institutions for these data to be interpreted).

5.2.3 Effects on One-Year Frequency of Reoffending (FR: 85 reports)

As with table 7, in each cell of table 8, we have presented the three figures consistently in this order: decrease/increase/inconclusive.

Table 8: Results for Frequency of Reoffending by Intervention and Sector

Intervention	Sector				
	Educational	Private	Public	VCS	Total
Accommodation			1/0/3	1/2/5	2/2/8
Arts				0/0/1	0/0/1
Education / Learning				7/0/1	7/0/1
Employment	1/0/4	7/0/13	6/0/11	3/0/3	17/0/31
Health and Wellbeing				0/0/1	0/0/1
Mentoring			0/0/1	0/0/5	0/0/6
Multi-Purpose				0/0/1	0/0/1
Problem Solving			0/0/1		0/0/1
Relationship Building				0/0/2	0/0/2
Restorative Justice					0/0/0
Substance Misuse				0/0/2	0/0/2
Youth Intervention			1/0/0	1/0/1	2/0/1
TOTAL	1/0/4	7/0/13	8/0/16	12/2/22	28/2/55

5.2.4 What Works in Reducing the Frequency of One-Year Reoffending?

Overall around 33% of interventions have been shown to reduce FR. Again more than half of these arise from within Employment interventions, but the success rate of Employment interventions is no greater than the overall rate. Success rates for Private, Public and Voluntary and Community Sectors are almost identical. Education and Learning interventions within the Voluntary and Community Sector appear particularly effective (7 decreases and 1 inconclusive).

5.2.5 Effects on Time to Reoffend (TR: 19 reports)

In table 9, the three numbers in each cell refer to the number of analyses which showed a significant increase in time to reoffend/a significant decrease in time to reoffend/or were inconclusive. Note that in order to facilitate comparison with tables 7 and 8, the increase and decrease labels have been reversed. Therefore the data below can be interpreted consistently with the results from PR and FR - that is, numbers in green (again the first ones presented) represent 'successful' results where it has taken longer for service users (the intervention group) to reoffend than for the comparison group. Thus, in table 9, each cell has figures presented in the following order: increase/decrease/inconclusive.

Table 9: Results for time to reoffend by Intervention and Sector

Intervention	Sector				
	Educational	Private	Public	VCS	Total
Accommodation				0/0/1	0/0/1
Arts					0/0/0
Education / Learning				5/0/3	5/0/3
Employment			0/0/1	0/0/1	0/0/2
Health and Wellbeing					0/0/0
Mentoring				0/0/3	0/0/3
Multi-Purpose				0/0/1	0/0/1
Problem Solving			0/0/1		0/0/1
Relationship Building				0/0/1	0/0/1
Restorative Justice					0/0/0
Substance Misuse				0/0/2	0/0/1
Youth Intervention				0/0/1	0/0/1
TOTAL	0/0/0	0/0/0	0/0/1	5/0/13	5/0/14

5.2.6 What Works in Increasing the Time to Reoffend?

The small number of reports (19) contained in the table above makes drawing any inferences problematic. The overall success rate is comparable with that observed for PR, and lower than for FR. It does appear that the effectiveness of Education and Learning interventions within the Voluntary and Community Sector is repeated when time to reoffend is used as the outcome variable.

5.3 Emerging Trends

There is some early aggregate evidence here that educational interventions run by the Prisoners' Education Trust and those focused on employment are more likely to demonstrate positive, statistically significant effects. However, it is also worth noting that in employment evaluations, more findings were inconclusive than were either positive, or negative. Also, in many sectors, there are still very few reports that have been prepared. The optimum size of sample presented to the JDL for analysis will be considered further below. For now, we consider in more depth, why it may be that there are so many gaps in the available data for aggregation.

The first and most obvious reason for gaps thus far, is that the JDL needs time to conduct extremely complex analyses and that its first 3 years of operation reflect the need to gain confidence, trust and "buy in" from public service provider organisations as well as the time needed to test, re-test and refine the model. Another reason may be that the provider organisations did not feel the need to use the JDL, as for example, they may not use

recidivism as their main outcome measure or those who commission their services may not require sight of JDL evaluations. Moving forward, it will be important to consider how successfully the JDL has reached out to potential clients and to consider how wide its potential client base may be. This may be particularly important given that the reports conducted so far have been completed for relatively few organisations (see Appendix 2). We therefore turn now to consider JDL client uptake and feedback.

6 Service Uptake and User Feedback

The JDL has conducted reviews of its processes, engagement and service user experiences. After the first year of operation, the JDL noted challenges in the quality, accuracy and quantity of data that organisations were able to provide. This had been recognised from the outset as a fundamental challenge for organisations and the JDL has developed and refined both how data are requested and the tools to support organisations in preparing and uploading data. Nonetheless it should first be noted that there has been wide variation in uptake rates both between different providers and different sectors (see tables 2 and 3). Also, across sectors, some intervention types appear to have been analysed much more frequently than others (educational interventions account for more than half the reports presented in table 3). In part, this is because of limitations to the initial model which meant that some types of intervention could not be included because of challenges in creating a comparison group. As the model has further developed and in particular, as OASYS data have been included, we would expect to see, for example, more analyses of substance misuse programmes.

In 2015[33], JDL compiled a user feedback survey based on 12 returns from the 34 organisations that had used the service up to that point. Organisations that had received statistically significant reports, were more likely to provide feedback than those where the findings had been inconclusive and they also tended to be positive in their feedback. The majority of users were very positive, saying that the reports had helped them to understand technical data, had been of use to inform trustees of their outcomes and to alert external agencies (including potential service commissioners), of their efficacy. The evidence had been used on websites, in reports and was generally taken as reliable, rigorous and helpful. Between March and September, 2015, NPC also conducted some research into JDL service uptake and feedback. Three hundred and eleven people were invited to take part in a survey about the JDL; they all worked in organisations with potential interest in using the JDL. Data

[33] Ministry of Justice, 2015b

were obtained from 53 different organisations, of which, 16 had used the service. Table 10 shows the sample characteristics, their familiarity with the service and some initial feedback.

Table 10: Sample Characteristics and Overview of Experiences

Familiarity with JDL (N=53)		User feedback (N=16)*		Non-user feedback (N=33)**	
Familiar and used JDL	16 (30%)	**Very positive**	4 (25%)	**Don't know enough**	12 (36%)
Familiar but have not used	23 (43%)	**Somewhat positive**	3 (19%)	**Not worked with enough service users**	8 (24%)
Not very familiar	10 (19%)	**Not positive**	4 (25%)	**Unable to get data**	8 (24%)
Never heard of it	4 (8%)	**Awaiting matches**	2 (12%)	**Service does not aim to reduce offending**	2 (6%)
		No comment made	3 (19%)	**Lack of consent**	2 (6%)
				Concerned that results will be published	2 (6%)
				Don't agree with methodology	1 (3%)
				Data security concerns	0 (0%)

**Some people selected more than one option.

These findings are not dissimilar to those from the JDL's feedback survey but reflect a slightly wider range of users. Those who had NOT used the service (whether familiar with it or not) were asked why they had not used it; table 10 indicates their responses to predefined categories. Free text answers to this question were also provided and included concerns about the timeliness or appropriateness of the outcome measures for the particular service users; that the type of intervention or service user was not eligible and that it was too resource intensive to collate and provide the data necessary. Examples included:

> *By the time data is available, the service is no longer in operation by us, and has changed significantly.*
>
> *It's difficult as we tend to work with longer term offenders so reduction in reoffending is difficult if they've not been released. It's too much of a risk to have it in the public eye and we'd need to get a very large samples*[34] *of data to make allowances for those not released.*

[34] Typographical errors included as typed by participants.

I think it was about the criteria - we don't collect all the necessary demographics of our service users that the data lab demanded. We needed more input/guidance on ways to get access to relevant data.

The above quotations all came from organisations reporting some familiarity with the JDL but no previous engagement with it. A more prosaic example from another organisation, also familiar with the JDL, shows that the intent may still be present, even if engagement hasn't happened:

just havent had time yet to do it

Those who had used the service (N=16) were asked to comment on how useful they found it. Four responses were positive, for example:

useful commissioned by MOJ - so great to receive a positive outcome on reducing reoffending rates

Good we plan to use in future

Three responses indicated that it had been somewhat useful, though with some limitations:

Limited use for our organisation. Working only with offenders in prison we need a more nuanced approach to assessing outcomes as the primary results of our work are intermediate outcomes on the reducing reoffending journey. The lack of subtlety in NOMS approach is frustrating and indeed damaging for our service users.

Our results were 'inconclusive', so we cannot say anything one way or another about the effectiveness of the intervention which was being analysed. However, it is useful to have engaged with Data Lab, if only to understand the complexities of it and its shortcomings. Our work has not been adversely affected by the 'inconclusive' outcome.

Two organisations were awaiting analyses/matches being found for their service users and the remaining 4 responses given identified problems such as inability to match sufficient clients, the currency or relevance of outcomes and not really finding inconclusive results useful:

Not particularly useful. Our main programme outcomes are not about reducing reoffending, they are about creating safer prisons, where prisoners can engage in rehabilitation

Not useful at all. At the time the historical data they had access to was not recent enough for us to have a large enough cohort to run any analysis.

Some of the limitations identified are built into the model and deliberately adopted, for example, the JDL focus is on interventions aiming to reduce recidivism (and who may be commissioned and paid by results in this domain). Thus organisations with intermediate outcomes which would not be expected to demonstrate change directly on recidivism, either need a different approach from the JDL entirely, or the JDL would need to consider whether proxy or intermediate indicators of recidivism could also be analysed. This would provide some help too, to programmes provided to longer serving prisoners, particularly, those which intervene early on in long sentences.

If ongoing support could be provided to offender intervention providers, it might be possible to help train and better facilitate data management and storage so that organisations would find it less taxing to provide the necessary data. If data storage and collation could be made more routine, then this would also help organisations when testing their models of change more generally and could be added value provided by the JDL.

7 Refining the Methodology and Reporting of Findings

The perceived rigour and reliability of the methodology employed by the JDL is at the heart of its credibility with academic and policy audiences, and methodology has been given considerable thought and attention by the JDL with key publications, at outset, one and at two year points of follow up[35]. An independent review of methodology was conducted in autumn, 2015 and the JDL responses to that review were published in spring, 2016[36].

Areas highlighted by the latest review of methodology include:

- minimum size of the intervention group;
- the start point of the 1 year proven reoffending rate;
- uses of other indicators of recidivism;
- better understanding of treatment needs and the incorporation of OASYS data;
- refining the matching process;
- data retention;
- testing for ‘unconfoundedness’.

[35] E,g, Ministry of Justice, 2016d
[36] Ministry of Justice, 2016a

For the current paper, we concentrate on the minimum size of the intervention group, the start point of the 1 year proven reoffending rate and consider data retention in the context of potential release of additional data into the public domain. We also consider how best to report findings and what additional data may be useful going forward.

7.1 Size of the Intervention Group and its Implications for Significance

The synthesis of key findings above highlighted a key challenge posed in running JDL analyses. A large part of the JDL's existence is to support organisations that would otherwise not be able to test for efficacy of their interventions, particularly when they have relatively low throughput of service users. Thus, the JDL has sought to keep the minimum numbers of service user (offender) data required relatively low. This has been aided by matching each intervention service user to many similar others in creating a comparison group. Nonetheless, the aggregate analyses clearly show that many more JDL reports are inconclusive than are statistically significant (in either positive or negative directions). This may be exacerbated by attrition rates as not all of the people included in the data provided to the JDL will remain in the intervention group eventually analysed. For example, people will be removed from the analysis when matches cannot be found for them[37].

In the review of methodology[38], it was suggested that the minimum size of cohort supplied to the JDL should be increased. This was rejected and in part, this may be to continue to provide support to smaller organisations, particularly to newer providers. However, the analyses included above indicate that the likelihood of gaining statistically significant results is much higher when there are more people put through in the intervention group.

It may be inappropriate to be asking for considerable input of time and effort in uploading data, particularly from smaller organisations, when inconclusive results are more likely than not. One way in which it might be inappropriate would be if the "inconclusive" findings are themselves deemed irrelevant. However, as noted in the user feedback survey, although some organisations do not know what to make of "inconclusive" findings, others found them useful. It might be possible to include an anonymised case study within a revamped user journey document that could show possible ways to use inconclusive findings, thereby encouraging the view of their likely occurrence in less negative ways.

[37] Some attrition may be expected when matching administrative data sets anyway, but additional failure to make a match may be for various reasons including that part of the cohort is either still in prison or has not been released for long enough to appear in reoffending data

[38] Ministry of Justice, 2016a

Another route to try to minimise the incidence of inconclusive findings would be to encourage more collaboration between organisations providing similar interventions to provide concurrent cohorts of service user data jointly to the JDL or to at least offer them the opportunity for their data to be analysed conjointly. This may be appropriate in situations where organisations have tried to use the service but have found that they have too few service users or where the analyses have led to inconclusive results. However, it is recognised that there are some external drivers that may make this problematic, for example, provider organisations may be in competition with one another or may be commissioned to provide services to different types of service user. Initiatives of this nature would need to be driven by the sector concerned, but could be supported or guided by the JDL in finding ways to account for the different implementation contexts within which organisations are delivering services.

It would also make sense to reappraise the information put out with “inconclusive” findings or where papers include “non-significant” results (see 7.4). Alongside greater use of effect sizes, significance could be posited as a threshold to be passed, not the ultimate arbiter of whether or not a report is meaningful. If this is done, then “inconclusive” findings could still be appraised in terms of the trends in the data towards positive or negative outcomes and the effect sizes could be considered on their own merits. This might enable the JDL to develop a system along the lines of “potentially promising” or “potentially worrying” flags that could be used alongside the more currently accepted “promising” and “effective” labels applied for policy decisions. If utilised, then this flagging system would be in keeping with other models but would still be distinct from WSIPP and NIJ in allowing for earlier consideration of findings (see section 3).

7.2 When Does Follow-Up Begin?

As part of this paper, we were asked to consider what overall syntheses could be conducted on JDL reports. We were mindful of the cautionary note that the JDL itself strikes in avoiding direct comparison between reports. This is because, although the model is standardised in its approach, it will not generate reports that are directly comparable.

One key problem in direct comparison stems from the framing of the one year follow up. All indices for reoffending used by JDL have been derived from the idea that follow up should be for one year. Whilst it is accepted that this is a relatively short time frame, particularly if trying to assess proven reconvictions for very serious crimes, it balances the need to follow up offenders in as timely a way as possible so as to feedback findings to the organisations running interventions and to trigger potential bonus payments based on positive results.

The starting point of the follow up year is conceptualised as being the point at which "offenders leave custody or start their sentence in the community"[39]. This is also referred to as the index date. For those who have served custodial sentences, it is the point of release into the community. This point has been taken for administrative consistency and efficiency[40]. As it currently stands, the index date introduces additional variance that militates against aggregating data. Not only do different interventions happen at different points in a sentence for different offenders, but the time lag after an intervention has been completed up to the point of release to the community will also vary. The time that a person spends in prison may be when they take other interventions and programmes, it may also be when they reoffend or even when prison management and post custody release policies are altered. The longer they spend in prison after completion of the intervention being assessed, the more the opportunity for additional positive and negative events. Thus the one year follow up is very rarely going to be one year after the actual intervention was completed and the longer that gap, the more likely it is that information is being missed. In order to capture more information about what happens after release, a time lag has been introduced, so that the 12 months of data are not collated until up to 18 months after release. The additional 6 months window, still may not allow for a completed court case when a serious re-offence occurs but does allow for more information to be collated about what happens post release, including potential ongoing supervision, of different extents and levels of intrusiveness.

If a sentence incorporates a community and custodial element, or is solely served in the community, then the 1 year follow up period commences with the start of the community based intervention. This means that offending that may occur during the intervention itself can be included in the JDL analysis for community sentences even though it is not currently included during custodial sentences. It also lowers the overall time available after an intervention has been completed. How much of the overall follow-up time is during the intervention and how much is after completion of the intervention will vary, depending on the length of the community sanction. A community order can last for up to 3 years, and it is possible that an intervention will run intermittently for the duration of that order. This means that the follow up period may be entirely subsumed within the implementation of the intervention. The 6 month time lag for data collection will again be somewhat useful but does not remove the need for caution when attempting to aggregate the JDL reports from across interventions.

[39] Ibid, p4
[40] We note that adjudications data were being considered for inclusion at the time of writing

7.3 Release of Data into the Public Domain

Data archiving has become an increasingly normative academic practice. Research Councils UK and many public sector research contracts require (suitably redacted) data to be archived and made publically available[41]. Similarly, journals are increasingly encouraging the free availability of anonymised data so that readers can run their own tests of the models being proposed in an article. Yet, the JDL functions on a clear policy of data destruction, not retention and is suitably mindful that they are not the data owners. This decision was taken to encourage organisations to come forward in the knowledge that their data could not subsequently be mined for information that they did not agree to/envisage originally.

The policy of not retaining data means that reanalyses of previous data sets cannot routinely be conducted as models are updated. Nor is it easy to tell which, if any, people who have offended have appeared more than once in a JDL analysis. Some organisations have worked with the JDL to develop protocols for keeping and reanalysing their data (e.g. the Prisoners Education Trust) but the normal route is that data will not be retained. This is also something that is being reviewed by the JDL, following the methodology review.

Even if data are retained, this does not mean that they can be made public. Firstly, the data are not owned by the JDL, and data publication would need to be with the permission of the data protection officer of the provider organisations (and in some cases, of the individuals who went through the intervention themselves, depending on what a priori consent had been obtained). If they are able to be passed forward, there would still be a need to protect the identity of the organisations submitting the data. Although data could be redacted and individual service users could be anonymised, the overall size of the datasets, points at which they were created and revised, etc, would mean that it would be relatively simple for someone who had read the published reports to “reverse engineer” the data and work out which dataset belonged to which organisation. If organisations (and their data controllers) are willing for this to happen then, going forward, it may be possible to archive the JDL datasets. This would allow for auditing, re-testing and for aggregation and synthesis, possibly via meta-analysis as more reports are run.

However, if organisations and service users wish to protect their data from general scrutiny, then the data might be archived with parameters in place to limit access, or the data may not be able to be publically archived at all. Instead, it might be possible to provide secure ways in which researchers could access the data, via the JDL itself. If this route were taken, we would expect that a framework would be consulted upon and agreed by the public sector

[41] See for example: http://www.rcuk.ac.uk/research/datapolicy/

provider organisations which have used the JDL that would articulate the circumstances and purposes for which the data could be used. These might include: to quality assure and allow the JDL process to be independently scrutinised; to facilitate external researchers or organisations themselves conducting analyses using indicators other than the recidivism measures currently employed or to conduct meta-analysis of a particular intervention type or sector.

7.4 Suggestions for JDL Reporting of Results

In JDL reports published concurrently with finalising the current paper, steps have been taken to make effect sizes more prominent and to begin to use more natural language. In particular, natural frequencies have been adopted in presenting differences between control and intervention groups[42]. This means that a reader can begin to interpret what appears to be happening, by reading side by side comparisons. We would suggest that this commendable start be further elaborated in subsequent papers and that natural frequencies be used to summarise differences, to provide a level of confidence in potential differences and to indicate some sensitivity analysis. All these matters are currently reported, but not in easy to interpret ways. This section posits some additional suggestions on reporting.

It is clear that establishing whether or not a particular intervention is effective is not straight forward, particularly where the data are incomplete, scarce or from a wide variety of sources. Even with complete, comparable data, there is 'no royal road to statistical induction'[43] and no single result answers the specific question we generally wish to answer - *does this work?*

The reality is that we are sometimes presented with an embarrassment of riches, a variety of results which individually may be suggestive, but collectively may be clearer or sometimes, more obscured. Here we propose how to present these different kinds of results in a meaningful way that should be accessible to readers with less statistical expertise, and supply three paving slabs for the road to induction.

1 What appears to be happening?[44]

Do the data suggest that the intervention reduces reoffending, and if so, by how much? Do the data suggest that the intervention increases reoffending, and if so, by how much? It is possible that there is no effect at all, but this is highly unlikely.

[42] For example, text box, page 1, August, 2016 Caritas report: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/544915/caritas-report.pdf
[43] Cohen, 1990, p1304
[44] Technically this is the effect size, which is not the same as the difference between two point estimates of prevalence, as is currently used. We recommend that JDL consider calculating formal effect sizes for all reports.

Data should be reported in meaningful formats which should not include percentages. Percentages can obscure the very information they are designed to convey. Instead, natural frequencies rounded up to a standard unit, are more intuitively meaningful e.g.

For every 100 people who receive this intervention, there will be x {fewer/more} crimes in the following year compared with similar people who do not receive the intervention.

2 How confident are we that this finding reflects the intervention's effect?[45]

How often might we see a result like this if there was no effect of the intervention but if the data collected arose for some other reason or set of reasons[46]?

If this intervention was actually ineffective, then we would expect to see this kind of difference around x% of the time. This implies that the intervention {is probably/might not be/is probably not}[47] causing the effect.

3 How accurate is this?[48]

We cannot be sure that the difference discussed in 1 above is the 'real' effect as we have only looked at a certain number of people. It is highly unlikely that we would get exactly the same difference if we measured a different group of people, but we can say with a high degree of confidence how different such an effect is likely to be from the original.

We are confident that were this analysis to be repeated with a new sample of the same size, the difference would be a {reduction/increase} of between x and y crimes in the following year.

Where this interval crosses zero, this would need to be further nuanced to indicate that additional caution should be exercised in interpreting the findings and these would be the kind of findings for which we would retain the label “inconclusive”.

[45] Technically, the traditional threshold for a test of significance, i.e. $p < .05$, currently included in reports and given a great deal more importance than here.
[46] Such as peculiarities of the individuals being treated, or delivering treatment, or the way in which data were recorded, or the creation of the control group against which comparisons were made.
[47] For threshold levels of significance, we suggest including a 'middle ground' for results which are promising (perhaps $p<0.2$).
[48] Technically this is the confidence interval, which is currently included in reports.

Integrating the results

We suggest that the three results be given equal weighting, but should be communicated in the order presented here. We further suggest adopting a summary framework along the lines of the WSIPP's, as outlined below.

The intervention appears to {reduce/increase}[49] *proven one year offending by x crimes for every one hundred people in receipt of the intervention. The data {allow/do not yet allow} us to assert this confidently, and we consider the intervention to be {effective/promising but unproven/concerning/inconclusive}*[50]*. While the reduction of x crimes is the best estimate based on the available evidence, we believe the real difference to lie somewhere between a {reduction/increase} of y and a {reduction/increase} of z crimes.*

7.5 Enhancing the Future Prospects for Synthesis

This paper began with a question about what a synthesis of initial JDL findings would tell us about criminal justice interventions. The first answer was that, it is not really possible to synthesise JDL findings. However, when a standard model is applied across a sector (e.g. youth justice), testing efficacy on a standard domain (e.g. one-year proven reoffending rate), the question of synthesis is tantalising for theoretical development, intervention improvement, policy outcome assessment and commissioning. We therefore end this section with consideration of what would be needed to enhance future prospects for synthesis.

Firstly, there are some basic statistical tests that are not routinely reported in the JDL reports. Whilst we do not wish to clutter the main report, it would seem sensible that the technical appendices include reports of means and standard deviations obtained. These would be necessary for any meta-analysis, for example.

To interpret future synthesis properly, it would also be necessary to know more about intervention processes. In particular, the duration, intensity and frequency of interventions are nowhere reported and no mention is made of how service users are referred to interventions. It was noted in section 5 that aggregating data about interventions within sectors was premature as so many cells were not populated at all. Asking for more data would seem to risk exacerbating this problem yet information about interventions, could be standardised across all sectors and all intervention types. As such, a database with this information, should be relatively quick to populate.

[49] From the effect size, as 1 above.
[50] From the significance test, with the 'concerning' label reserved for effects which are significant in the 'wrong' direction.

8 Summary of Findings and Recommendations

Before presenting our overall conclusions and recommendations it is important to note that the driver for this paper was a request for synthesis and review. At this stage in the progress and development of the JDL, synthesis may have been premature. We would hope that as uptake increases and a wider variety of providers and interventions are analysed, trends will start to emerge and patterns in reoffending reduction may be discerned.

At the moment, we would summarise the key findings thus:

1. The JDL has been generally well received by those who have used it, they find the information useful and several have chosen to use the service repeatedly.
2. It is encouraging to note that so few negative results have been found, this indicates a generally healthy set of interventions running in our criminal and youth justice systems.
3. We note however that the majority of findings have been labelled as inconclusive and uncertainty about how to use such findings may be limiting uptake or reuse of the service.
4. Effect sizes did not vary with the size of the intervention group analysed. In other words, whether a small or large cohort was put through the analysis, the magnitude of change observed did not vary significantly. This is a positive finding and indicates both that change *can* be observed and that it can be discerned despite the undoubted noise in the model. However, it is not surprising that the changes observed are failing to make statistical significance when observed within small sample sizes.
5. It is hard to discern many trends given the relatively limited uptake of the service and challenges faced in aggregating the published data (see section 3). However, educational interventions seem to be repeatedly showing as being effective (and are largely provided by the Prisoner Educational Trust) and there are some positive findings coming out of employability/employment initiatives. Within the employment interventions however, there is more variability in outcome and this would warrant further investigation. These types of intervention are also shown to be effective within WSIPP data where specific educational and employment initiatives have been shown as being certain to generate positive returns on investment (see Appendix 1).

At this stage, we think that there is positive evidence that the JDL has made a solid start to its operations. The staff have worked hard to refine the statistical modelling and continue to strive for excellence in producing rigorous, reliable evidence. However, uptake has been relatively low. Where organisations have engaged with the process, some have been able to test efficacy, refine their interventions and demonstrate outcomes with some statistically reliable evidence behind them. For others, the statistical evidence has been more equivocal

and they have found it harder to utilise. This is a manifest challenge that has been acknowledged several times in reviews but has not been sufficiently addressed. There is scope here for the expert panel and for the JDL to consider in some depth how best to proceed in dealing with inconclusive findings.

The JDL can contribute to policy decisions and where results are statistically significant, it is clear that commissioning decisions (including payments for results obtained) can be directly informed by the analyses. Here too, it is harder to make this case with the current reporting provided for inconclusive findings. The model is necessarily predicated on a design that looks for a relatively weak signal across quite a noisy field, and as such, effect sizes were never likely to be particularly high. Where effect sizes are relatively modest (albeit consistent and frequently in desirable directions), cohort size is likely to have a substantial influence on whether or not statistically significant results can be obtained.

8.1 Recommendations

Effect sizes should be more prominent in reports and natural language should be used to communicate findings. Although confidence intervals are reported, effect sizes seem to have been marginal to the reporting model adopted. We note that they will be incorporated more prominently going forward and hope that they will be used in ways that explain the magnitude and implications of observed changes. This should go alongside a clear statement that statistical significance is a threshold to be obtained and not the ultimate indicator of efficacy. This can be done in accessible ways that will allow data from interventions currently labelled as inconclusive to be drawn on, limiting those labelled as inconclusive only to interventions where findings seem to pull in different directions (and are thus quite realistically depicted as inconclusive). It may also be worth considering benchmarking effect sizes that can be expected for a particular intervention in a particular sector. This could be done where meta-analyses already exist, or using reappraisals of the WSIPP data (such as that conducted by the Dartington Social Research Unit (2012/3). Effect sizes could then be compared against the benchmarks.

Increase uptake and engagement. The JDL reports are easy to find, once one knows that they exist. If not aware of them already, one is unlikely to find them. The reports are not indexed and there is little prospect of this happening. They could perhaps be set up in ways that would improve search engine hits, possibly with key words to include “outcome evaluation” or “recidivism rates”. The JDL needs to be more present, not just at statistical seminars, but in practitioner events and its online footprint needs to be larger. It is also worth considering whether previous JDL service users could be invited to act as ambassadors for the service, possibly creating vox pops/vlogs and otherwise raising the profile of the service

amongst those who have not engaged with it. It is also important to note that the JDL model will not always be the most effective or appropriate means of analysis. If trust and engagement are to be effective, it will be as important to highlight when not to use the service as when to engage with the JDL.

Where organisations have tried to use the JDL service but have found that they have too few service users or where the analyses have led to inconclusive results, efforts could be increased to find ways for different organisations providing similar interventions to have their data collated and analysed together. This would need to be sector driven and although guidance from the JDL would also be needed, given the small size of the JDL team and their workload, this might be an excellent area for the expert panel to support.

Provide more support for using the findings. The expert panel and JDL could also be drawn on further to help organisations to do more with the findings generated. At the moment, the “what can you say” panels are clear and provide neat summaries, but they are not an appropriate place to make suggestions about where to use findings and how to say something about the findings, targeted to the different audiences that might be interested in them. Better support for use of the findings would also help enhance engagement.

Set up a means to retain and make data available. We concur with suggestions to retain data and note that data security has never been identified as a problem in user feedback surveys (even when offered as a pre-defined choice). We agree with Research Councils UK that data are a public good and that secondary analyses could be invaluable. Whether archived publically or not, is something that would need careful consideration but we hope that data retention will come to replace post analysis data destruction as the normative practice within JDL protocols.

Collate more information on intervention practices. In particular, information about intervention frequency, duration, intensity and referral would be useful. If reported alongside means and standard deviations, this would better enable the prospect for meta-analysis and other systematic forms of synthesis.

9 References

Adler, J. R; Edwards, S. K; Scally, M.; Gill, D.; Puniskis, M. J.; Gekoski, A; & Horvath, M. A. H. (2016) *What Works in Managing Young People who Offend? A Summary of the International Evidence.* Ministry of Justice Analytical Services https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/498493/what-works-in-managing-young-people-who-offend.pdf

Aos, S. & Drake, E. (2013). Prison, police, and programs: Evidence-based options that reduce crime and save money. (Doc. No. 13-11-1901). Olympia: Washington State Institute for Public Policy.

Cohen, J. (1990). Things I Have Learned (So Far). *American Psychologist.* 45 (12) 1304-1312. http://www.stats.org.uk/statistical-inference/Cohen1990.pdf

Davis, Lois M., Robert Bozick, Jennifer L. Steele, Jessica Saunders and Jeremy N. V. Miles. Evaluating the Effectiveness of Correctional Education: A Meta-Analysis of Programs That Provide Education to Incarcerated Adults. Santa Monica, CA: RAND Corporation, 2013. http://www.rand.org/pubs/research_reports/RR266.html.

Dowden, C., & Andrews, D. (2004). The importance of staff practice in delivering effective correctional treatment: A meta-analytic review of core correctional practice. International Journal of Offender Therapy and Comparative Criminology, 48, 203-214. doi:10.1177/0306624X03257765

England, R. (2009). Adolescents with sexually inappropriate behaviours: Background, Characteristics and approaches to intervention. (Unpublished PhD). University of Liverpool

Garrido, V. & Morales, L.A. (2007). Serious (violent and chronic) juvenile offenders: a Systematic Review of treatment effectiveness in secure corrections. Campbell Systematic Reviews, 7, 3. 1-46. doi:10.4073/csr.2007.7

Haines, K., Case, S, Davies, K., & Anthony, C. (2013) The Swansea Bureau: A model of diversion from the Youth Justice System. International Journal of Law, Crime and Justice, 1-21 http://dx.doi.org/10.1016/j.ijlcj.2013.04.002

Koehler, J. A., Hamilton, L., & Lösel, F. A. (2013a). Correctional treatment programmes for young offenders in Europe: A survey of routine practice. European Journal on Criminal Policy and Research, 19, 387-400. doi:http://dx.doi.org/10.1007/s10610-013-9206-6

Koehler, J. A., Lösel, F., Akoensi, T. D.,& Humphreys, D. K. (2013b). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. Journal of Experimental Criminology, 9, 19–43. doi:10.1007/s11292-012-9159-7

Larkins, C., & Wainwright, J. (2013). 'Just putting me on the right track' young people's perspectives on what helps them stop offending. Preston; University of Central Lancashire. Retrieved from http://clok.uclan.ac.uk/9483/1/Larkins%20and%20Wainwright%202014%20-%20right%20track%20PUBLISHED.pdf

Lipsey, M. W., Wilson, D. B., & Cothern, L. (2000). Effective intervention for serious juvenile offenders. Washington, USA: US Department of Justice.

Lipton, D.S., Martinson, R. and Wilks, J. (1975) The effectiveness of correctional treatment: A survey of treatment valuation studies. New York, NY: Praeger Press.

Martinson, R. (1974). What Works? - Questions and Answers About Prison Reform. The Public Interest, 35, 22-54.

Mears, D. P., Cochran, J. C., Greenman, S. J., Bhati, A. S., & Greenwald, M. A. (2011). Evidence on the effectiveness of juvenile court sanctions. Journal of Criminal Justice, 39, 509-520. doi:10.1016/j.jcrimjus.2011.09.006.

Miller, M., Fumia, D., & He, L. (2015). The King County Education and Employment Training (EET) Program: Outcome evaluation and benefit-cost analysis. (Doc. No. 15-12-3901). Olympia: Washington State Institute for Public Policy.

Ministry of Justice (2016a). *Justice Data Lab: A Peer Review of existing methodology – Response* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/506327/methodology-review-response.pdf

Ministry of Justice (March, 2016b). *Justice Data Lab Statistics, May2016* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/522589/justice-data-lab-statistics-may-2016.pdf

Ministry of Justice (March, 2016c) *Justice Data Lab Publication Summary, March 2016* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/522440/jdl-publication-summary-may-2016.xlsx

Ministry of Justice (2016d). *Justice Data Lab: Incorporating Offender Assessment data to the Justice Data Lab process – Methodology* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/491688/oasys-methodology.pdf

Ministry of Justice (June, 2015a). *Justice Data Lab. Pilot summary* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/434165/justice-data-lab-pilot-summary.pdf

Ministry of Justice (2015b). *Justice Data Lab: What has the pilot told us? User feedback summary*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/434167/justice-data-lab-feedback-report.pdf

Ministry of Justice (2014a) *Justice Data Lab User Journey Document*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/392927/justice-data-lab-user-journey.pdf

Ministry of Justice (2014b). *Justice Data Lab: The pilot year.* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/289250/justice-data-lab-pilot-year.pdf

Ministry of Justice (April, 2013). *Justice Data Lab Methodology Paper* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/392929/justice-data-lab-methodology.pdf

Ministry of Justice. (2012/13). *Compendium of Re-offending Statistics and Analysis*. Ministry of Justice Statistics Bulletin. HMSO: Justice Statistics Analytical Services. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/278133/compendium-reoffending-stats-2013.pdf

Petrosino A., Turpin-Petrosino C.,& Guckenburg, S. (2010) Formal system processing of juveniles: Effects on delinquency. Campbell Systematic Reviews, 1, 3-80. doi:10.4073/csr.2010.1

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M., Lavenberg, J. G. (2013). Scared straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review. Campbell Systematic Reviews, 5, 6-53. doi:10.4073/csr.2013.5

Redondo, S., SanchezMeca, J., & Garrido, V. (1999). The influence of treatment programmes on the recidivism of juvenile and adult offenders: An European meta-analytic review. Psychology, Crime & Law, 5, 251-278. doi: 10.1080/10683169908401769

Redondo, S., SanchezMeca, J., & Garrido, V. (1999). The influence of treatment programmes on the recidivism of juvenile and adult offenders: An European meta-analytic review. Psychology, Crime & Law, 5, 251-278. doi: 10.1080/10683169908401769.

Rickey, B. & Pritchard, D. (2012). *Unlocking Offender Data. How access to offending data could help charities improve outcomes for offenders* London. NPC http://www.thinknpc.org/publications/unlocking-offending-data/

Ross, A., Duckworth, K., Smith, D. J., Wyness, G., & Schoon, I. (2011). Prevention and reduction: A review of strategies for intervening early to prevent or reduce youth crime and anti-social behaviour. (DFE-RR111). London, England: Department for Education.

Schwalbe, C. S., Gearing, R. E., MacKenzie, M. J., Brewer, K. B., & Ibrahim, R. (2012). A meta-analysis of experimental studies of diversion programs for juvenile offenders. Clinical Psychology Review, 32, 26-33. doi:10.1016/j.cpr.2011.10.002

Small, S.A., Reynolds, A.J., O'Connor, C., & Cooney, S.M. (2005). What Works, Wisconsin: What science tells us about cost-effective programs for juvenile delinquency prevention. Madison, WI: University of Wisconsin–Madison.

Smith, E.P., Wolf, A.M., Cantillon, D.M., Thomas, O., & Davidson, W.S. (2004). The adolescent diversion project: 25 years of research on an ecological model of intervention. Journal of Prevention & Intervention in the Community, 27(2), 29-47

Social Research Unit. (2012). Youth Justice, Investing in Children. Dartington: Social Research Unit.

Social Research Unit. (2013). Investing in Children; Technical Report. Dartington: Social Research Unit.

Strang, H., Sherman, L. W., Mayo-Wilson, E., Woods, D., & Ariel, B. (2013). *Restorative Justice Conferencing (RJC) Using Face-to-Face Meetings of Offenders and Victims: Effects on Offender Recidivism and Victim Satisfaction. A Systematic Review*. Campbell Systematic Reviews, 9(12).

10 Appendix 1: Summary of WSIPP Findings

Table A1 Juvenile Justice Interventions Reviewed by WSIPP and likelihood of positive return on investment

Program name	Chance benefits will exceed costs
Wilderness experience programs for juvenile offenders	**100%**
Adolescent Diversion Project	**100%**
Education and Employment Training (EET, King County)	**100%**
Functional Family Therapy (youth in state institutions)	**99%**
Functional Family Therapy (youth on probation)	**99%**
Family-based therapy (Parenting with Love and Limits model)	**98%**
Diversion, no services (vs. traditional juvenile court processing)	**97%**
Diversion with services (vs. traditional juvenile court processing)	**97%**
Coordination of Services	**96%**
Other family-based therapies (non-name brand)	**95%**
Aggression Replacement Training (youth in state institutions)	**93%**
Aggression Replacement Training (youth on probation)	**92%**
Mentoring	**88%**
Victim offender mediation	**80%**
Multisystemic Therapy	**78%**
Therapeutic communities for chemically dependent juvenile offenders	**78%**
Cognitive Behavioral Therapy (CBT) for juvenile offenders	**78%**
Functional Family Parole (with quality assurance)	**73%**
Family Integrated Transitions (youth in state institutions)	**69%**
Group homes (Teaching-Family Model)	**67%**
Multidimensional Treatment Foster Care	**63%**
Drug court	**58%**
Vocational and employment training for juvenile offenders	**54%**
Multisystemic Therapy for substance abusing juvenile offenders	**54%**
Intensive supervision (parole)	**50%**
Diversion with services (vs. simple release)	**40%**
Other chemical dependency treatment for juveniles (non-therapeutic communities)	**23%**
Scared Straight	**4%**
Intensive supervision (probation)	**0%**

Table A2 Adult Criminal Justice Interventions Reviewed by WSIPP and likelihood of positive return on investment

Program name	Chance benefits will exceed costs
Swift and Certain sanctions for offenders on community supervision	**100%**
Drug courts	**100%**
Electronic monitoring (parole)	**100%**
Vocational education in prison	**100%**
Correctional education (basic or post-secondary) in prison	**100%**
Risk Need & Responsivity supervision (for high and moderate risk offenders)	**100%**
Cognitive behavioural treatment (for high and moderate risk offenders)	**100%**
Inpatient/intensive outpatient drug treatment (incarceration)	**100%**
Therapeutic communities for chemically dependent offenders (community)	**100%**
Outpatient/non-intensive drug treatment (incarceration)	**100%**
Correctional industries in prison	**100%**
Employment & job training assistance in the community	**100%**
Mental health courts	**100%**
Therapeutic communities for offenders with co-occuring disorders	**99%**
Employment & job training assistance during incarceration	**99%**
Work release	**98%**
Drug Offender Sentencing Alternative (for drug offenders)	**98%**
Case management: swift & certain/graduated sanctions for substance abusing offenders	**95%**
Therapeutic communities for chemically dependent offenders (incarceration)	**94%**
Electronic monitoring (probation)	**94%**
Offender Re-entry Community Safety Program (dangerously mentally ill offenders)	**94%**
Sex offender treatment in the community	**93%**
Day reporting centers	**92%**
Outpatient/non-intensive drug treatment (community)	**90%**
Sex offender treatment during incarceration	**75%**
Intensive supervision (surveillance & treatment)	**72%**
Restorative justice conferencing	**71%**
Drug Offender Sentencing Alternative (for property offenders)	**70%**
Jail diversion programs for offenders with mental illness (post-arrest programs)	**57%**
Inpatient/intensive outpatient drug treatment (community)	**51%**
Case management: not swift and certain for substance abusing offenders	**44%**
Domestic violence perpetrator treatment (Duluth-based model)	**17%**
Intensive supervision (surveillance only)	**6%**
Prison	
For lower risk offenders, decrease prison average daily population by 250, by lowering length of stay by 3 months	**71 %**
For moderate risk offenders, decrease prison average daily population by 250, by lowering length of stay by 3 months	**10 %**

For high risk offenders, decrease prison average daily population by 250, by lowering length of stay by 3 months	**0 %**
Police (results per-officer)	
Deploy one additional police officer with hot spots strategies	**100 %**
Deploy one additional police officer with State wide average practices	**100 %**

Data downloaded from: http://www.wsipp.wa.gov/BenefitCost?topicId=2 accessed 23.6.16

11 Appendix 2: Assumptions Underpinning the Synthesis

Samples from which individual JDL reports are generated are not completely independent, and some recipients of interventions are double (or even triple) counted. This arises for several reasons:

1. For any one intervention, regional analyses are aggregated in order to produce a national analysis. This means that service users have been analysed at least twice: the first time involves using a comparison sample drawn from the region where recipients originate, and subsequently with a national comparison sample;

2. Repeat analyses have been conducted for several organisations both in order to test the evolving JDL model and to allow organisations to re-run analyses as more of their users have gone through the programme or intervention and as additional years of comparative data have been added to the JDL database. Unless a specific organisation has agreed to allow JDL to retain their data then JDL's normal data retention policy makes it impossible to determine the extent to which recipients in subsequent cohorts may overlap with those in the originals;

3. It is possible that the same people may have engaged in more than one intervention that has been analysed. Also, that individuals included in a comparison group for one analysis may have been included in the intervention group for another. For individual analyses, this is not particularly problematic but it should be borne in mind when data are aggregated across analyses.

4. Lastly, it should be noted that different programmes may be designed to run with different types of service user. Even if they are both focussed on the same skill or outcome, such as literacy or employability skills, if aimed to run with different client groups, then the individual propensity matching will be amended accordingly. Such differences cannot so easily be considered when aggregating data and the JDL published reports are not designed to be directly compared with one another.

Thus, before aggregating any data (as we have been asked to do), it is critical to determine which reports should be considered and we need to be clear about the assumptions we have made. Our two core assumptions are that:

1 National comparison samples are superior to regional comparison groups as they result in larger matched samples, and correspondingly greater statistical power. In short, using a national comparison sample is *a priori* more likely to detect an effect, should one exist.

2 Second requests are more reliable indicators of an intervention's effect than first requests. Even if sample sizes are no greater for second requests, and do not result in greater or smaller effects, the dynamic and ongoing nature of the model employed to conduct JDL analyses means that more recent results are likely to be more accurate estimates of interventions' efficacy.

We therefore synthesise only reports that are national and, where more than one request has been made, we draw on the most recent iteration.

11.1 Organisations included in the 97 reports analysed.

1. A4e - First Steps Programme
2. Adelaide House - Approved Premise
3. Blue Sky - short term, full-time employment contracts
4. Brighton & Hove City Council - Preventing Offender Accommodation Loss (POAL) Project
5. Community Justice Court (CJC) at Plymouth Magistrates' Court
6. DISC - Leeds Drug Intervention Programme/Integrated Offender Management programme (Leeds DIP/IOM project)
7. Everyday Skills
8. Foundation
9. GOALS UK
10. HMP Downview D Wing Resettlement Unit
11. HMP Kirklevington Grange
12. HMP Swansea Community Chaplaincy Project
13. Home Group Residential and support service - Delivered after prison sentences
14. Inside Out (formerly known as Wormwood Scrubs Community Chaplaincy)
15. Lancashire Women's Centres
16. Langley House Trust
17. Leap
18. NOMS Bail Accommodation and Support Services (BASS) - Bail with a conditional discharge or fine
19. Nottingham Women's Centre
20. Only Connect
21. Phoenix Futures - Therapeutic Communities Programme
22. Pre-school Learning Alliance
23. Prince's Trust
24. Prison Fellowship - Sycamore Tree
25. Prisoners Education Trust - Overall analyses - second request
26. Riverside ECHG
27. Roundabout
28. Safe Ground - Family Man programme - fifth request
29. Shelter housing advice / assessment sessions in HMP Leeds
30. St. Helens Integrated Offender Management
31. The Footprints Project
32. The Koestler Trust - Koestler Trust awards
33. The Prison Phoenix Trust
34. Time for Families - first request
35. Warwickshire Youth Justice Service
36. West Yorkshire Community Chaplaincy Project
37. Women's Centres throughout England
38. Working Chance - first request